
Analyse des patrimoines de données géographiques nationaux.

Comparaison de trois infrastructures nationales de données géographiques (France, Brésil, Bolivie)

Julie Pierson¹, Matthieu Noucher¹, Pierre Gautreau², Louca Lerch³, Olivier Pissotat¹, Aurély Jautard¹, Sylvain Lesage⁴

1. UMR 5185 ADESS, CNRS et Université de Bordeaux
Maison des Suds, 12 Esplanade des Antilles, F-33607 Pessac Cedex, France
j.pierson@ades.cnrs.fr ; m.noucher@ades.cnrs.fr ; o.pissotat@ades.cnrs.fr ;
a.jautard@gmail.com
2. UMR 8586 PRODIG, Université Paris 1 Panthéon Sorbonne
2 rue Valette, F-75005 Paris, France.
pierre.gautreau@univ-paris1.fr
3. Université de Genève, Département de Géographie et Environnement
Uni Mail, 40 Bd du Pont-d'Arve, CH-1211 Genève 4, Suisse
louca.lerch@unige.ch
4. GeoBolivia - Vicepresidencia del Estado Plurinacional de Bolivia
(jusqu'en janvier 2013)
severo@rednegra.net

RÉSUMÉ. Par le double effet de l'interopérabilité des systèmes et de l'évolution du cadre légal, les patrimoines de données géographiques institutionnelles tendent aujourd'hui à être de plus en plus facilement accessibles à travers la mise en place généralisée d'infrastructures de données géographiques (IDG). L'analyse comparée du contenu de trois IDG nationales (France, Bolivie, Brésil) permet alors de révéler quelques tendances. L'originalité de cette contribution est donc de faire des IDG non pas un support à la recherche d'information mais un véritable objet de recherche. L'analyse des métadonnées des trois catalogues nationaux nous permettra alors de réinterroger les logiques de formalisation et de diffusion des connaissances sur les territoires.

ABSTRACT. Due to the dual impact of systems interoperability and changes in the legal framework, institutional spatial data tends to become more and more accessible through widespread deployment of spatial data infrastructures (SDI). Comparative analysis of the content of three national SDI (France, Bolivia, Brazil) reveals a few tendencies. This contribution's originality lies in using SDI not as an aid to research but as a true object of study. Metadata analysis of three national catalogs offers a way to reconsider the logics of both formalisation and knowledge diffusion through territories.

MOTS-CLÉS : infrastructure de données géographiques ; métadonnée ; géocatalogue ; thésaurus ; emprise géographique ; ISO 19115.

KEYWORDS: spatial data infrastructure ; metadata ; geocatalog ; thesaurus ; bounding box ; ISO 19115.

1. Introduction

La diffusion croissante d'outils, de données, de méthodes et la mise en réseau des SIG rendent aujourd'hui leur analyse incontournable pour comprendre la constitution des politiques publiques en matière d'environnement, en particulier la genèse de certaines normes en matière de représentations et d'évaluation territoriale. Ainsi, le SIG devient un point de passage obligé (d'Alessandro-Scarpari et *al.*, 2008) (Amelot, 2012). Celui-ci tend aujourd'hui à se renforcer avec, sous le fait des évolutions législatives (illustrées par la mise en œuvre – en cours et jusqu'en 2020 – d'un « arsenal juridique » autour de la Directive INSPIRE) et de l'accroissement des capacités techniques (liées notamment aux efforts de normalisation), une diffusion exponentielle de flux de données, en particulier de données géographiques (Noucher et Gautreau, 2013).

Dans ce contexte, les infrastructures de données géographiques qui visent à favoriser la mise en réseau des SIG, en facilitant l'accessibilité aux données géographiques, semblent un objet de recherche intéressant pour comprendre, par l'analyse de leur contenu, les logiques de formalisation et de diffusion des connaissances sur les territoires. Nous proposons ici de revenir rapidement sur la notion d'IDG (chap. 2) puis de présenter les trois IDG nationales analysées (chap. 3) avant d'indiquer les éléments d'analyse des métadonnées qui alimenteront le poster présenté à l'occasion de la conférence SAGEO'13 (chap. 4).

2. Les infrastructures de données géographiques au centre des dynamiques de diffusion des connaissances territoriales institutionnelles.

Une infrastructure de données géographiques (IDG) a pour but de permettre aux utilisateurs un accès direct à des informations et services géographiques de différents fournisseurs. Une IDG se compose ainsi d'un ensemble de technologies, de politiques, de normes et de ressources humaines pour acquérir, traiter, stocker, distribuer et améliorer l'utilisation de l'information géographique (Rajabifard et *al.*, 2003) (Crompvoets et *al.*, 2004). Les IDG existent à différents niveaux (Masser, 2010) selon le principe de subsidiarité. Ainsi, en Europe l'infrastructure européenne de données géographiques repose sur des infrastructures nationales (en France, d'un point de vue technique, le niveau national est représenté par le couple 'géoportail IGN / géocatalogue BRGM') qui elles-mêmes reposent sur des IDG régionales (par exemple, PIGMA en Aquitaine), départementales (par exemple, IGECOM 40 dans les Landes) voire locales.

Ces IDG nous semblent aujourd'hui intéressantes à analyser pour, *a minima*, deux raisons. Premièrement, le mouvement de structuration et de diffusion de l'information géographique institutionnelle au sein des IDG a été initié depuis une dizaine d'années. Par conséquent, on dispose désormais d'un historique

potentiellement intéressant à considérer (analyse rétrospective de la constitution des bases de métadonnées, des emprises disponibles, des échelles de travail, des nomenclatures...). Deuxièmement, malgré l'apparition progressive des normes en vigueur (et des guides de bonnes pratiques qui y sont associées), les stades d'avancement sont encore très différents. Les IDG travaillent donc avec des approches et des priorités thématiques différentes selon les territoires et permettent de comparer des compréhensions / appropriations différentes des normes en cours de constitution.

3. La comparaison des INDG française, brésilienne et bolivienne pour réinterroger les données géographiques disponibles dans des contextes d'usage différents.

Selon les degrés de pénétration des techniques web et géomatiques, des niveaux de conflictualités et des pratiques de décision autour des données géographiques, pour une même thématique les productions de données peuvent radicalement différer. L'étude de contextes sociétaux différents permet donc d'ouvrir une perspective comparative apte à réinterroger les patrimoines de données géographiques disponibles dans les différents systèmes étudiés. Dans cette logique, nous avons étudié trois catalogues de données géographiques issus d'infrastructures nationales du Brésil, de Bolivie et de France. Le choix d'étudier simultanément et de façon comparée les usages sociaux des IDG dans un pays européen (la France) et des pays sud-américains (Bolivie, Brésil) doit permettre en particulier de repérer une grande diversité de bases, du fait d'histoires institutionnelles fortement contrastées, d'identifier des modes divers de circulation de l'information environnementale (des formes les plus institutionnalisées aux plus informelles selon un gradient France-Brésil-Bolivie). Ce choix est également lié à la disponibilité dans ces trois pays de catalogues de données géographiques nationaux dont la structuration des métadonnées repose, en partie, sur la même norme.

Pour chacun de ces catalogues, les fiches de métadonnées au format XML ont été extraites au moyen d'un script en Python créé pour l'occasion. Cette extraction est programmée pour se répéter automatiquement tous les trois mois, la première extraction ayant eu lieu au 1er avril 2013. Les fichiers XML extraits répondent aux normes internationales ISO 19115 (définition du contenu des métadonnées) et ISO 19139 (définition de la structure des métadonnées).

Tableau 1: Nombre de fichiers XML extraits par géocatalogue au 01/04/2013.

| Pays | Géocatalogue | Nombre de fichiers XML |
|-------------|---|-------------------------------|
| Bolivie | http://www.geo.gob.bo/catalogapp/ | 527 |
| Brésil | http://www.inde.gov.br/ | 8466 |
| France | http://www.geocatalogue.fr/ | 35303 |

On remarque d'emblée une très forte disparité dans le volume de ces catalogues (cf. Tableau 1) qui s'explique à la fois par des historiques différents (le géocatalogue français date de 2007 alors que le géocatalogue bolivien a été lancé en 2012) mais aussi par des modalités de saisie différentes et des organisations territoriales plus ou moins complexes. Par exemple, sur les 35 303 fiches du géocatalogue français, on compte plus de 23 000 fiches produites par la DGFIP¹ et qui concernent le cadastre à l'échelon communal, une couverture continue à très grande échelle sur l'ensemble du territoire. A l'autre extrême, le catalogue de l'IDG bolivienne révèle de grandes disparités en termes d'information disponible suivant que l'on se trouve sur des territoires privilégiés par les acteurs internationaux (coopération inter- et non-gouvernementale, multinationales) ou non. L'analyse des métadonnées ne peut donc se contenter d'un comptage des fiches pour essayer de comprendre des logiques informationnelles des pays étudiés.

4. L'analyse des métadonnées comme révélateur des logiques de formation et de diffusion des connaissances territoriales.

L'analyse du contenu des champs de métadonnées doit permettre de disposer d'une information plus fine susceptible d'alimenter l'approche comparative préconisée. Pour ce faire, des requêtes en langage XQuery permettent d'interroger les 90 balises XML identifiées comme similaires entre les trois catalogues. Trois axes d'analyse, dont les résultats seront présentés sous la forme de cartes et/ou de graphiques statistiques sur le poster, sont privilégiés :

Où : La couverture territoriale des données géographiques est-elle homogène ?

Pour répondre à cette question, nous analysons les emprises géographiques définies dans les fiches de métadonnées. L'extraction des 4 coordonnées des rectangles d'emprise (un même jeu de données peut posséder plusieurs emprises si la couverture est discontinue) est réalisée à partir d'un script en Python qui permet de récupérer des champs d'un catalogue CSW, d'en extraire les balises XML correspondantes aux emprises (<EX_GeographicBoundingBox>) en format CSV pour générer ensuite une couche SIG en format SHP. Un traitement dans un logiciel SIG, permet enfin, pour chaque pays, de produire une carte des densités de données géographiques. Cette *géographie de l'information géographique* révèle alors des répartitions hétérogènes de l'information et des discontinuités spatiales qui ne sont pas le seul fait de disparité démographique. Ainsi, en Bolivie par exemple, l'aire urbaine de Cochabamba apparaît comme beaucoup moins couverte que des secteurs quasi désertiques comme l'Oruro ou le Potosí. L'analyse de la couverture géographique des données cartographiques institutionnelles permet donc de révéler des situations potentiellement intéressantes à approfondir pour comprendre les dynamiques territoriales et/ou politiques de constitution des patrimoines de données géographiques.

¹Direction Générale des Finances Publiques

Quoi : Les thématiques prioritaires sont-elles les mêmes ?

L'analyse des thésaurus et surtout la comparaison des catégories affectées aux 40.000 jeux de données des trois catalogues nationaux étudiés permet d'identifier les thématiques prioritaires et de comparer leur répartition. Cette tâche s'est révélée complexe dans la mesure où les trois catalogues ne respectent que partiellement la norme ISO19915. Il a donc été nécessaire de concevoir une table d'appariement qui assure une jointure entre des méta-catégories et les catégories utilisées dans chaque catalogue national. Ainsi le hiatus entre les catégories INSPIRE et les catégories ISO a, par exemple, nécessité quelques regroupements. Une fois cet appariement réalisé la comparaison entre pays permet de *donner à voir* les préoccupations prioritaires. L'analyse dans le temps de ces comparaisons permettra, à terme, d'identifier l'évolution des thématiques et de révéler ainsi la transformation des questionnements.

Quand : La constitution des patrimoines s'opère-t-elle au même rythme ?

Enfin, l'analyse des informations temporelles permet d'identifier à la fois les grandes phases de constitution des patrimoines de données et de mises en partage. La datation de la constitution peut se faire par l'analyse de trois balises XML relatives à des champs de métadonnées qui concernent la date de création et/ou la date de révision et/ou la date de publication de la donnée. Cependant, comme l'illustre le tableau 2 la saisie de ces trois types de dates est très inégale d'une fiche de métadonnées à l'autre et d'un catalogue à l'autre. Dans ces conditions il devient complexe de ne sélectionner qu'un champ de métadonnées pour comparer les trois catalogues.

Tableau 2 : part des métadonnées renseignée par type de 'date'.

| | France | Bolivie | Brésil |
|---|---------------|----------------|---------------|
| % de métadonnées avec date de création renseignée | 90 | 4 | 54 |
| % de métadonnées avec date de publication renseignée | 26 | 71 | 33 |
| % de métadonnées avec date de révision renseignée | 20 | 8 | 1 |

Par ailleurs, l'analyse des grandes phases de mise en partage des données peut potentiellement s'opérer à partir des dates de création des métadonnées. Ces dates sont saisies automatiquement à chaque création de fiche. La comparaison sur des trois géocatalogues est donc possible. Elle révèle des rythmes de montées en charge des catalogues différents (progressif au Brésil, par palier en France et en Bolivie). Les ruptures observées peuvent alors s'expliquer par les évolutions du cadre réglementaire (par exemple, la publication du règlement INSPIRE concernant les métadonnées le 3 décembre 2008 marque le début d'une montée en charge du

géocatalogue français) ou par des logiques organisationnelles (mise en place de d'ateliers de saisie de métadonnées en Bolivie à partir de mi-2012).

5. Conclusion

L'objectif des extractions et analyses en cours est de pouvoir caractériser l'information présente dans les catalogues de trois pays (Brésil, France, Bolivie), de manière reproductible dans le temps. Au-delà des schémas et cartes issus de ce travail qui seront présentés sur le poster l'intérêt de ce travail est aussi de montrer les difficultés méthodologiques associées à l'évaluation des infrastructures de données géographiques pour l'analyse des patrimoines de données géographiques institutionnelles. Les écarts aux normes internationales se révèlent alors être à la fois une difficulté technique pour engager la comparaison de catalogues différents et une opportunité pour analyser les interprétations divergentes et adaptations locales qui révèlent la variété des dynamiques organisationnelles et territoriales sous-jacentes à ces dispositifs socio-techniques.

Remerciements:

Cette contribution s'inscrit dans le cadre du projet de recherche BAGUALA dont l'objectif est d'analyser les impacts de la diffusion des données environnementales en libre accès sur les pratiques de gestion de l'environnement. Pour plus d'information : <http://baguala.hypotheses.org>

Bibliographie

- Amelot X. (2012). Cartographie participative et web 2.0 pour le développement local et la gestion de l'environnement à Madagascar, empowerment ou impérialisme numérique ?, *Séminaire GRANIT « Cartographie et participation : quand la cartographie critique et la cartographie 2.0 se rencontrent. »*, UMR ADESS, Pessac.
- Crompvoets J., Bregt A., Rajabifard A., Williamson I. (2004). Assessing the worldwide developments of national spatial data clearinghouses. *International Journal of Geographical Information Science*, vol. 18, p. 665-689.
- D'Alessandro-Scarpari C., Elmes G., Weiner D. (2008). L'impérialisme numérique. Une réflexion sur les Peace Parks en Afrique australe, *Géocarrefour*, vol. 83, n°1, p. 35-44.
- Masser I. (2010). *Building European Spatial Data Infrastructures, Second Edition*, ESRI Press, Redlands, California.
- Noucher M., Gautreau P. (2013). Le libre accès rebat les cartes. *Les Cahiers du Numérique*, vol.9, n°1, p. 57-83.
- Rajabifard A., Feeney M., Williamson I.P., Masser I. (2003). Chapter 6, National SDI Initiatives. *Development of Spatial Data Infrastructures: from Concept to Reality*. London, Taylor & Francis, p. 95-109.